



Open Archive TOULOUSE Archive Ouverte (OATAO)

OATAO is an open access repository that collects the work of Toulouse researchers and makes it freely available over the web where possible.

This is an author-deposited version published in : <http://oatao.univ-toulouse.fr/>
Eprints ID : 17090

To link to this article : DOI : 10.3166/is.21.2.65-94
URL : <http://dx.doi.org/10.3166/is.21.2.65-94>

To cite this version : Canut, Marie-Françoise and On-At, Sirinya and Péninou, André and Sèdes, Florence *Construction du profil social de l'utilisateur dans un contexte dynamique : application d'une méthode de pondération temporelle*. (2016) Ingénierie des Systèmes d'Information, vol. 21 (n° 2). pp. 65-94. ISSN 1633-1311

Any correspondence concerning this service should be sent to the repository administrator: staff-oatao@listes-diff.inp-toulouse.fr

Construction du profil social de l'utilisateur dans un contexte dynamique

Application d'une méthode de pondération temporelle

**Marie-Françoise Canut, Sirinya On-at, André Péninou,
Florence Sèdes**

*Institut de Recherche en Informatique de Toulouse (IRIT)
Université de Toulouse, CNRS, INPT, UPS, UT1, UT2J, 31062 Toulouse Cedex 9
{marie-francoise.canut; sirinya.on-at; andre.peninou; florence.sedes}@irit.fr*

RÉSUMÉ. Les mécanismes d'adaptation de l'information (personnalisation, recommandation) s'appuient tous sur un profil utilisateur. Une des difficultés majeures est la construction de ce profil dont la pertinence vis-à-vis des besoins/intérêts de l'utilisateur joue un rôle important dans la qualité des adaptations produites. Les réseaux sociaux numériques (RSN) sont une source d'informations très riche sur l'utilisateur. Nous nous intéressons au processus de construction du profil utilisateur à partir de son RSN appelé « profil social ». Nous proposons dans ce travail d'intégrer des critères temporels dans le processus de construction du profil de l'utilisateur. Le poids « temporel », qui reflète la pertinence d'un intérêt à un instant donné, est calculé, d'une part, à partir de la pertinence des individus du réseau égocentrique de l'utilisateur en prenant en compte « la fraîcheur » de leurs liens avec l'utilisateur et, d'autre part, à partir de la pertinence des informations qu'ils partagent en prenant en compte la « fraîcheur » de ces informations. Cet article est une version étendue de celui présenté à Inforsid 2015 (Canut et al., 2015) en proposant un état de l'art plus complet, la description et l'analyse d'expérimentations plus poussées.

ABSTRACT. Adaptive mechanisms (personalization, recommendation) all rely on user profile. A major issue is the construction of such profile. Its relevance regarding needs and interests of the user plays an important role in the quality of adaptation. Online social networks provide valuable information sources about the user. This work focuses on extracting user interests from his/her social network to build his/her social profile. We propose to apply a time-aware method into an existing egocentric-based user profiling process. The aim of this strategy is to weight user's interests according to its relevance and freshness. The time awareness weight of an interest is computed by combining the relevance of individuals in the user's egocentric network, computed by taking into account the freshness of their ties, and the information relevance, computed by taking into account its freshness. This paper is an extended version of our previous work presented at Inforsid2015 (Canut et al., 2015). We propose a more detailed study of the state of the art and further experiments and analysis.

MOTS-CLÉS : profil utilisateur, réseau social, réseau égocentrique, pondération temporelle.

KEYWORDS: user profile, social network, egocentric network, time-weighted method.

1. Introduction

Pour pouvoir offrir des informations qui correspondent aux besoins de l'utilisateur, les mécanismes de personnalisation/adaptation doivent disposer d'informations sur les utilisateurs telles que leurs caractéristiques personnelles, leurs préférences, leurs centres d'intérêt, communément appelées profil utilisateur. Une des difficultés majeures est la construction de ce profil dont la pertinence vis-à-vis des besoins/intérêts de l'utilisateur joue un rôle important dans la qualité des adaptations produites. De ce fait, le profil utilisateur devient central dans les systèmes d'adaptation de l'information. Dans la littérature, plusieurs modèles de profil utilisateur ont été proposés (Gauch *et al.*, 2007). Nous représentons le profil utilisateur sous forme d'un vecteur de termes pondérés qui décrivent les centres d'intérêt de l'utilisateur définis donc par un ensemble (vecteur) de mots clés (foot, tennis, danse, etc.) pondérés.

Les réseaux sociaux numériques (RSN) de plus en plus utilisés, apparaissent comme une source riche d'information sur les utilisateurs. (Tchunte, 2013) a proposé un processus de construction du profil utilisateur à partir de son réseau social représenté par un graphe de relations entre les individus. Ce processus permet d'une part, de compléter le profil d'un nouvel utilisateur ou celui d'un utilisateur peu actif pour lequel le profil serait insuffisant pour les mécanismes d'adaptation, et d'autre part, d'enrichir un profil existant. Nous utilisons le terme profil social pour désigner un profil construit par ce type de processus.

Toutefois, ces travaux ne prennent pas en compte le caractère dynamique des RSN. En effet, un RSN est considéré comme un réseau évolutif car le comportement de ses utilisateurs (reflétant leurs intérêts) évolue de manière particulièrement rapide (mobilité des utilisateurs et développement de nouveaux outils de connexion et de partage d'informations). Cet aspect pourrait entraîner une obsolescence rapide de certaines relations ou de certaines informations partagées et donc amener à une mauvaise interprétation de leurs intérêts en l'absence de tout filtrage de ces données. Nous nous trouvons face à la problématique suivante : comment prendre en compte la caractéristique évolutive et dynamique des RSN pour construire un profil social à la fois pertinent et à jour ?

Pour répondre à ce problème, nous nous appuyons sur les travaux de (Tchunte *et al.*, 2013) et nous proposons d'utiliser des critères temporels pour sélectionner efficacement les informations provenant de ce réseau. Nous envisageons donc dans ce travail, de construire le profil social de l'utilisateur en intégrant une mesure temporelle (reflétant la pertinence) dans le calcul du poids des intérêts liés au profil social de l'utilisateur. Ce poids, appelé poids temporel, est calculé à partir, d'une part, des individus en lien avec l'utilisateur et d'autre part, des informations que ces individus partagent.

Ce travail fait suite à la proposition présentée à Inforsid 2015 (Canut *et al.*, 2015) avec un état de l'art et un positionnement des travaux plus précis ainsi qu'une description et une analyse d'expérimentations plus détaillées.

Nous présentons tout d'abord le contexte de notre travail sur la construction du profil social, puis nous insistons sur la caractéristique dynamique des RSN et la gestion de l'évolution des intérêts dans le profil social construit. Nous présentons ensuite la prise en compte de critères temporels dans le processus de construction du profil social de l'utilisateur. Ensuite, nous décrivons et analysons les différentes expérimentations effectuées dans deux réseaux de publications scientifiques (DBLP et Mendeley). Nous terminons par une conclusion et les perspectives de ce travail.

2. Contexte du travail

Généralement, le profil utilisateur peut être construit, d'une part à partir des activités propres à l'utilisateur (ceci peut être fait de façon explicite par l'utilisateur ou de façon implicite par l'extraction des intérêts de l'utilisateur lors de ses interactions avec le système), d'autre part en utilisant des sources externes comme des utilisateurs similaires (largement utilisées dans le cas de filtrage collaboratif) ou bien à partir des membres dans son réseau social (Abel *et al.*, 2011 ; Gauch *et al.*, 2007; Mezghani *et al.*, 2012; Tchuente *et al.*, 2013).

Nous utilisons dans ce papier, le terme « profil social » pour désigner un profil construit par ces approches. Le profil social de l'utilisateur (où les intérêts de l'utilisateur sont enrichis à partir des informations partagées par les individus de son réseau social) permet de compléter la représentation du profil de l'utilisateur mais aussi de résoudre le problème de démarrage à froid ou du cas d'un utilisateur très peu actif (qui interagit très peu avec le système et fournit donc moins d'informations pour déduire ses intérêts) (cf. figure 1). Le profil social de l'utilisateur peut être construit soit en se basant sur des utilisateurs considérés de manière individuelle (Cabanac, 2011 ; Carmel *et al.*, 2009), soit en se basant sur les communautés de son réseau social (Tchuente *et al.*, 2013).

Cabanac (Cabanac, 2011) s'appuie principalement sur des utilisateurs considérés individuellement pour proposer un système de recommandation sociale d'articles scientifiques aux chercheurs. (Carmel *et al.*, 2009) utilise le même principe pour proposer un système de recherche d'information sociale. Ces méthodes peuvent être définies comme des méthodes de type « autoritaire » pour lesquelles les utilisateurs les plus actifs ou les plus influents dans le réseau social seront privilégiés dans la dérivation du profil social d'un autre utilisateur. Pourtant, dans le contexte des RSN, il est difficile de considérer que tous les centres d'intérêt d'un utilisateur influent dans le réseau social, peuvent représenter fidèlement cet utilisateur. Ce problème peut être mieux perçu si l'on considère des environnements de RSN comme Facebook dans lesquels un utilisateur peut être ami avec plus de 1 000 individus. Parmi ces individus, très peu sont réellement ses amis proches qui fournissent des informations permettant de caractériser correctement cet utilisateur, même s'ils sont très actifs ou influents.

Les travaux de (Tchuenté *et al.*, 2013) portent sur l'étude de la dérivation du profil utilisateur à partir des communautés de son réseau égocentrique, un réseau largement utilisé en sociologie. Il s'agit d'un graphe composé des relations entre les individus situés à distance 1 (directement reliés) de l'utilisateur (appelé *égo*), l'*égo* étant bien entendu exclu de ce graphe. Ces travaux proposent donc une méthode plutôt de type « affinitaire », dans laquelle c'est la présence d'affinités, de liens, de relations entre les individus d'une communauté du réseau égocentrique de l'utilisateur qui permet de dériver des informations à associer à son profil. Ainsi, le profil est dérivé à partir des communautés (calculées) regroupant les individus du réseau égocentrique et non plus des individus pris séparément. Par rapport aux méthodes autoritaires, cette approche éliminera, d'une part, les individus ayant des relations superflues¹ avec l'utilisateur (ces individus, sans lien avec le réseau social de l'utilisateur, n'appartiendront à aucune communauté), et d'autre part, les éléments de profil non significatifs pour l'utilisateur (pertinence des éléments de profils liés à l'affinité des liens). Dans ce travail, nous nous intéressons à cette dernière approche, en considérant que les utilisateurs qui créent des liens entre eux sont ceux qui partagent les mêmes points communs (intérêts, géographie, établissement, etc.) (Aral et Walker, 2013).



Figure 1. Profil utilisateur versus profil social

2.1. Le processus de dérivation du profil utilisateur à partir de son réseau égocentrique (CoBSP)

CoBSP (Community Based Social Profile) propose un modèle générique du profil utilisateur suivant deux dimensions : une dimension utilisateur dont les centres d'intérêt sont calculés à partir des activités propres de l'utilisateur et une dimension

¹ Amitiés acceptées au hasard ou pour démontrer un certain pouvoir par le nombre d'amis par exemple.

sociale (profil social) dont les centres d'intérêt sont calculés à partir des activités de son réseau égocentrique. Ces deux dimensions étant complémentaires et indépendantes, elles peuvent être utilisées par les mécanismes d'adaptation individuellement ou couplées. L'étude présente ensuite un processus de dérivation du profil social de l'utilisateur à partir des communautés extraites de son réseau égocentrique. Nous expliquons brièvement le processus CoBSP en 4 étapes successives décrites ci-après :

1) La première étape consiste à extraire, depuis le réseau égocentrique d'un utilisateur, les communautés de ce réseau en utilisant l'algorithme iLCD proposé par (Cazabet *et al.*, 2012). Cet algorithme se base sur la structure du réseau pour extraire des communautés et prend en compte le recouvrement des communautés.

2) La deuxième étape consiste à calculer le profil de chaque communauté détectée dans la première étape (vecteur de termes/éléments). Le profil d'une communauté peut être calculé en utilisant les informations de tous les individus qui en font partie (ici, les poids associés aux éléments présents dans la dimension utilisateur de chaque individu).

3) La troisième étape consiste à pondérer le profil de chaque communauté en se basant sur une caractérisation structurelle et/ou une caractérisation sémantique. La caractérisation structurelle d'une communauté se base sur la centralité de degré (*degree centrality*). Selon cette mesure, la communauté qui possède le plus grand nombre de connexions directes dans le réseau est caractérisée comme étant la plus importante par rapport aux autres communautés. La caractérisation sémantique d'un élément de profil d'une communauté consiste à rechercher sa spécificité par rapport aux autres communautés en se basant sur la mesure de pondération tf-idf entre les intérêts des communautés. Les deux caractérisations sont combinées pour obtenir une caractérisation unique (sémantico-structurelle). La caractérisation sémantico-structurelle de chaque élément e du profil d'une communauté C_i sera calculée par la formule suivante :

$$\text{Caractérisation finale}(e) = \alpha \cdot \text{Struct}(C_i) + (1-\alpha) \cdot \text{Sem}(C_i) \quad (1)$$

Le paramètre α (valeur dans l'intervalle $[0,1]$) dans la formule permettra de juger et de faire varier l'importance des mesures de structure dans la dérivation du profil social de l'utilisateur par rapport aux mesures sémantiques. La valeur de α est déterminée de manière empirique et sera donc fixée lors des expérimentations.

4) La quatrième étape permet de dériver les intérêts du profil social par combinaison des différents poids associés à un intérêt à partir de toutes les communautés en utilisant une fonction de combinaison linéaire classique (Shaw *et al.*, 1994).

2.2. Problème et motivation : intérêts de l'utilisateur et problème de dérive conceptuelle (*concept drift*)

La gestion dynamique du profil utilisateur est un problème soulevé et non traité dans le travail existant de (Tchuente *et al.*, 2013). En effet, les centres d'intérêt d'une personne sont amenés à évoluer dans le temps (changement d'environnement ou de contexte du travail). Le problème de l'évolution des intérêts des utilisateurs peut être considéré comme celui d'une dérive conceptuelle « *concept drift* » généré par la caractéristique évolutive des concepts au cours de temps (Gama *et al.*, 2014; Tsymbal, 2004 ; Widmer et Kubat, 1993). Ce type de dérive peut se produire soudainement (*sudden concept drift*) ou bien graduellement (*gradual concept drift*) alors qualifiée (selon le taux de changement de concepts) de dérive modérée ou de dérive lente (Stanley, 2003). La cause d'un changement de concept peut être soit autonome, soit liée à un phénomène externe (changement dans son environnement) appelée alors dérivation conceptuelle virtuelle (*virtual concept drift*) (Widmer et Kubat, 1993). Dans notre contexte de travail, nous nous intéressons tout particulièrement à la dérivation conceptuelle virtuelle : comme les intérêts du profil social de l'utilisateur sont extraits à partir des informations de son réseau social, le changement ou l'évolution de ce dernier impacte de fait, l'évolution des intérêts du profil social. Cette évolution reflète la dérivation conceptuelle virtuelle des intérêts de l'utilisateur.

Nous partons du constat que les intérêts de l'utilisateur évoluent au fil du temps, en particulier, dans le cas d'intérêts extraits depuis les RSN. En effet, dans de tels réseaux, l'information partagée évolue sans cesse du fait des interactions sociales en ligne (partage, échange d'informations) qui génèrent (plus facilement) un volume important d'informations volatiles. Les intérêts qui sont extraits à une période donnée peuvent ne plus être significatifs ultérieurement. Par exemple, pour un utilisateur qui regarde le foot uniquement pendant la période de la coupe du monde (une fois tous les 4 ans) et qui partage des informations dans son réseau social, l'intérêt pour la coupe du monde ou le foot n'est significatif pour son profil que pendant cette période.

Un autre problème qui peut être rencontré est la pertinence des liens de l'utilisateur avec les membres de son RSN ainsi que celle des informations qu'il partage. Les utilisateurs peuvent créer des contacts en ligne sans forcément connaître les personnes dans la vie réelle. Si on reprend l'exemple précédent, un utilisateur peut suivre les informations à partir des comptes de joueurs de foot qui sont alors enregistrés comme contacts sans pour autant les connaître personnellement. Après la compétition, les liens avec ces joueurs deviennent moins importants pour l'utilisateur. Cela nous montre bien que l'on ne peut pas prendre en compte (ou donner la même importance à) toutes les informations existant dans les RSN pour refléter les intérêts d'un utilisateur à un moment donné.

Nous nous trouvons donc face à la problématique suivante : dans le contexte des RSN, comment construire un profil social de l'utilisateur qui soit à la fois pertinent

et à jour ? En s'appuyant sur la proposition de (Tchuenté, 2013), nos travaux tentent de prendre en compte la gestion dynamique du profil utilisateur dans le processus de dérivation du profil social. Il nous paraît alors important d'étudier l'évolution de ce profil en fonction de l'évolution du RSN de cet utilisateur.

3. État de l'art

3.1. Évolution du réseau social

Pendant plusieurs années, l'étude des propriétés et des caractéristiques des réseaux sociaux (densité, degré de distribution, classification, composants connexes, communautés) a été considérée comme une piste importante de recherche. Cependant, la plupart des études ont été conduites avec une vision globale et statique du réseau alors qu'un RSN est considéré comme un réseau dynamique (qui évolue au fil du temps). Pour répondre à ce problème, l'analyse de la dynamique du réseau social ne peut s'effectuer qu'en prenant en compte un critère temporel pour comprendre les évolutions qui se produisent dans le réseau (Aggarwal et Subbian, 2014). Nous pouvons distinguer deux types de dynamique dans un réseau social : la dynamique de la structure du réseau et la dynamique des informations partagées.

3.1.1. Dynamique de la structure du réseau social (dynamique dans le réseau)

La dynamique de la structure du réseau social provient de la dynamique des interactions des individus dans le réseau. Elle est liée à la création et/ou à la suppression des liens mais aussi à la persistance des liens déjà créés.

Plusieurs travaux portent sur l'évolution au niveau global (macroscopique) (Kumar *et al.*, 2006 ; Lin *et al.*, 2009). (Kumar *et al.*, 2006) ont étudié l'évolution de la structure des RSN (Flickr² et Yahoo! 360³) et ont constaté une évolution similaire des deux réseaux qui se caractérisent par une croissance rapide des relations et des informations, suivie d'une diminution puis une reprise de croissance lente mais régulière. Suite à leurs analyses, ils ont pu identifier trois groupes d'utilisateurs : 1) les « singletons », qui participent au service mais ne participent pas à l'activité du réseau social (pas d'interaction ni de connexion avec les autres utilisateurs), 2) les « linkers » qui se trouvent dans le groupe « Giant Component » comprenant les utilisateurs les plus actifs et sociables. Ils se connectent soit *via* des liens (invitations), soit directement par la recherche d'amis *via* la plateforme, 3) les utilisateurs du groupe « *Middle Region* » qui représentent les petites communautés isolées qui n'ont que des interactions internes au groupe (et pas avec les utilisateurs des autres groupes). Cette étude a prouvé qu'avec le temps, les groupes « *Middle Region* » ont tendance à s'unir avec des groupes « Giant Component ».

2. www.flickr.com

3. www.360.yahoo.com (clôturé depuis juillet 2009).

Plusieurs travaux portent sur l'évolution au niveau local (microscopique) de la dynamique des liens entre utilisateurs (Leskovec *et al.*, 2010 ; Rivera *et al.*, 2010). Selon (Rivera *et al.*, 2010), nous pouvons distinguer différents mécanismes fondamentaux qui peuvent générer la dynamique des liens (création). Le mécanisme « *TriadicClosure* » se base sur le principe d'amis en commun (deux individus indépendants qui ont des amis en commun ont une grande probabilité de devenir amis). Le mécanisme « *Preferential Attachment* » s'applique par rapport au fait qu'un individu a tendance à se connecter avec d'autres individus qui ont déjà beaucoup de connexions (populaires) dans leurs réseaux (degré élevé). Le mécanisme « *Homophily* » met en œuvre le fait que des personnes ayant les mêmes goûts ont plus de chances de se connecter entre elles. Par exemple, des chercheurs travaillant sur la même thématique ont tendance à collaborer et à échanger des informations. Le mécanisme « *Global Connection* » s'appuie sur le principe que l'individu peut créer des contacts au-delà de ses réseaux ou de ses proches. Par exemple, les professionnels se contactent pour collaborer et créent des nouveaux réseaux professionnels. Et enfin, le mécanisme « *Random* » se fonde sur le fait que l'individu peut se connecter avec d'autres individus de façon totalement aléatoire.

On retrouve l'utilisation de ces mécanismes dans les travaux de prédiction de liens (Liben-Nowell et Kleinberg, 2003) qui utilisent ces différents mécanismes s'appuyant sur la topologie de réseau (nombre de voisins communs, le plus court chemin, attachement préférentiel, etc.) pour prédire la formation de liens entre les nœuds dans le réseau.

3.1.2. Dynamique des informations dans le réseau (dynamique sur le réseau)

Le deuxième facteur clé est lié à l'analyse temporelle des informations dans le réseau social et porte sur le partage et la diffusion d'informations. L'analyse de la dynamique des informations porte sur le modèle de diffusion d'informations, sur la recherche des nœuds influents dans le réseau mais aussi sur les techniques permettant de diffuser efficacement des informations dans le réseau social (Jiang *et al.*, 2014). D'après plusieurs travaux de recherche, la structure du réseau a un impact important sur la dynamique des informations dans le réseau. Les individus qui sont considérés proches dans le réseau ont une probabilité plus grande de s'échanger des informations. Dans le cas inverse, la dynamique de la diffusion d'informations peut également être à son tour, un facteur pour le changement de la structure du réseau social (Stattner *et al.*, 2013 ; Weng *et al.*, 2013).

3.1.3. Dynamique du comportement des utilisateurs dans le contexte des RSN

L'apparition des RSN facilite l'évolution rapide des réseaux sociaux sous-jacents en termes de structure de réseau mais aussi de flux d'informations (Gomez Rodriguez *et al.*, 2013). En termes de structure de réseau, les mécanismes comme « *TriadicClosure* » ou « *Homophily* » gérant la dynamique de la structure du réseau social peuvent se produire plus facilement et rapidement dans les réseaux sociaux en ligne. Étant donné que les liens peuvent être établis facilement, deux utilisateurs

reliés entre eux ne sont pas obligés de se connaître dans la réalité. De ce fait, la persistance des liens n'est pas toujours maintenue. Dans (Arnaboldi *et al.*, 2013) les auteurs ont analysé le réseau social Twitter⁴ pour étudier comment les relations sociales en ligne sont maintenues. Ce travail a montré que les utilisateurs ont tendance à avoir des relations faibles, et il y a beaucoup de « *turnover* » dans les contacts de leurs réseaux. Ainsi, pour un utilisateur, ses relations peuvent évoluer au fil du temps. Le lien entre deux utilisateurs peut être pertinent pendant une certaine durée et peut devenir de moins en moins significatif par la suite.

En termes de flux d'informations, le mécanisme de partage d'informations peut être le fruit d'événements sociaux ou d'un marketing viral (*buzz*) qui favorise le partage de contenus en ligne. Souvent, ce phénomène de société arrive pendant une période courte puis disparaît progressivement et peut réapparaître à une autre période. Ainsi, ce phénomène peut mener à une explosion d'informations dans les réseaux sociaux en ligne, par exemple, la guerre civile libyenne, la révolution égyptienne de 2011 (Rodriguez *et al.*, 2012). En outre, il montre que le comportement de partage d'informations dans un réseau social a une caractéristique dynamique et non persistante. Certaines informations partagées par un utilisateur peuvent ne pas toujours refléter son comportement sur le long terme.

Plusieurs travaux sur l'évolution du réseau social ne considèrent pas le changement d'informations dans le réseau comme un facteur important pour l'analyse du réseau (Stattner *et al.*, 2013). Toutefois, dans le contexte des RSN dont les membres créent et partagent une masse d'informations importante, les données circulant dans le réseau deviennent volatiles et évoluent rapidement. Les facteurs comme les *buzz* ou les campagnes marketing peuvent favoriser la diffusion d'informations dans les RSN mais aussi la création de nouveaux liens entre les personnes qui s'intéressent à un événement. Les techniques existantes de traitement des informations doivent être adaptées aux caractéristiques de ces nouvelles sources d'informations afin d'obtenir des informations les plus représentatives de l'utilisateur. Il est donc important de prendre en compte non seulement la dynamique de la structure du réseau mais également la dynamique des informations dans l'analyse de l'évolution du réseau.

Ainsi, les caractéristiques dynamiques des RSN nous amènent à considérer qu'il n'est pas certain que les individus du réseau social de l'utilisateur ainsi que les informations qu'ils partagent à un instant t soient pertinents et à jour à l'instant $t + \Delta t$. Notre proposition s'attache à prendre en compte ce constat dans le processus de dérivation du profil social de l'utilisateur.

4. www.twitter.com

3.2. Gestion de l'évolution des intérêts dans le profil utilisateur

L'étude de l'évolution des intérêts de l'utilisateur consiste à prendre en compte leurs changements à travers le temps (Crabtree *et al.*, 1998). Nous pouvons distinguer deux approches de gestion de l'évolution des intérêts dans le profil utilisateur. La première approche consiste à gérer la dynamique des intérêts de l'utilisateur après la phase d'extraction des intérêts et correspond à un processus de mise à jour du profil utilisateur (Mianowska et Nguyen, 2013 ; Sugiyama *et al.*, 2004). La seconde approche consiste à prendre en compte la dynamique des centres d'intérêt pendant l'étape d'extraction des intérêts. Nous nous intéressons dans ce travail à la seconde approche, dans le but de construire un profil social pertinent dès sa première utilisation.

Comme indiqué dans la section 2.2, l'évolution des intérêts peut être considérée comme un problème de dérivation conceptuelle. Pour traiter ce problème nous pouvons appliquer les techniques décrites par la suite et définies par trois grandes approches selon (Koren, 2009 ; Tsybal, 2004).

La première est l'approche par sélection d'instance (*instance selection*) qui choisit des informations qui sont pertinentes par rapport à une période choisie. Ce raisonnement a été utilisé dans la construction du profil utilisateur dans plusieurs contextes. Plusieurs travaux proposent d'utiliser les informations à court terme pour mieux cerner l'évolution et la dynamique des informations étudiées. La plupart de ces travaux se basent sur l'approche « *time-forgotten* » qui ignore les informations trop anciennes (Maloof et Michalski, 2000). Dans la recherche d'information personnalisée, on utilise l'historique à court terme de l'utilisateur lié à une seule (la dernière) session de recherche pour extraire ses intérêts (Bennett *et al.*, 2012). Les techniques utilisées dans l'approche de sélection d'instance oublient complètement les informations dépassant une date limite. Pourtant, certaines informations ignorées peuvent s'avérer pertinentes et ne pas les prendre en compte peut entraîner une perte d'informations intéressantes. En effet, (Tan *et al.*, 2006) ont prouvé que l'historique de recherche à long terme est très important pour améliorer la tâche de recherche d'informations dans le cas de requêtes récurrentes.

La deuxième approche appelée approche pondérée (*instance weighting*) calcule le poids de chaque instance selon son poids de pertinence estimé. Souvent c'est une fonction temporelle (*time decay function*) qui est utilisée pour donner plus de poids aux informations les plus récentes. Dans ce type d'approche, toutes les informations existantes de l'utilisateur peuvent être exploitées mais de manière plus restreinte. Ce raisonnement a été utilisé dans plusieurs travaux sur les systèmes de recommandation (Li *et al.*, 2013, 2014). Cette idée peut être retrouvée également dans le contexte de la recherche d'information personnalisée comme dans (Kacem *et al.*, 2014) qui proposent d'appliquer une fonction temporelle pour pondérer les intérêts de l'utilisateur selon leur fraîcheur.

La troisième approche est basée sur l'apprentissage en utilisant plusieurs prédicteurs pour calculer les résultats de prédiction (*ensemble learning*). Dans un système de recommandation de films, (Koren, 2009) propose d'extraire le modèle de l'évolution du comportement d'utilisateurs dans tout l'intervalle de temps en utilisant plusieurs prédicteurs. Les évaluations réalisées par (Koren, 2009) montrent l'efficacité de cette approche par rapport aux deux autres dans le contexte des systèmes de recommandation. Cependant cette approche nécessite de disposer de données d'apprentissage pour obtenir des résultats pertinents.

3.3. Synthèse

Il est important de souligner que dans nos travaux, les principales interactions que nous souhaitons étudier pour détecter un changement de centre d'intérêt ne sont pas ciblées sur l'utilisateur lui-même mais sur les éléments de son réseau social (liens entre les membres, informations qui circulent entre les membres) : l'évolution du profil social de l'utilisateur est liée à l'évolution de son réseau social. Pour étudier l'évolution des intérêts de l'utilisateur dans notre contexte, il est donc important de prendre en compte des critères temporels en s'appuyant sur les mécanismes d'évolution du réseau social de l'utilisateur.

En se basant sur la caractéristique évolutive des informations ou des relations dans un réseau social numérique, certains facteurs peuvent avoir un impact sur le comportement des utilisateurs. Les utilisateurs actifs (*Giant Component*) ont tendance à partager beaucoup d'informations et créent des liens pendant une période donnée. Il existe également des utilisateurs passifs (*Singleton, Middle Region*) dont les réseaux sociaux évoluent lentement. Le réseau social de l'utilisateur peut donc évoluer de manière progressive ou bien soudainement lors d'événements donnant lieu à une accélération des partages. Pour construire un profil social pertinent, il faut choisir une technique qui évite les simples accumulations d'informations et permet d'en valoriser certaines et d'en atténuer d'autres. La technique d'apprentissage avec plusieurs familles de prédicteurs semble être la méthode la plus pertinente mais elle nécessite des données d'apprentissage, ce qui ne pourra pas vraiment être possible certains types d'utilisateurs en particulier pour les nouveaux utilisateurs ou les utilisateurs peu actifs. La technique de « *instance selection* » présente un risque d'oubli d'information, et ne permet pas la valorisation des informations importantes. Si on utilise les informations sur le long terme pour construire un profil, on risque d'avoir des données accumulées obsolètes et donc mal interprétées.

Nous nous intéressons donc à la technique « *instance weighting* » qui permet de sélectionner et traiter des informations de manière restreinte en prenant en compte des critères temporels. Une fois le profil construit de manière pertinente, nous pouvons penser maintenir sa pertinence en appliquant des techniques de mise à jour de profil utilisateur. Les approches de sélection d'instances ou l'approche d'apprentissage deviendront sans doute plus adaptées dans ce cas, puisqu'elles pourront se baser sur un profil déjà construit. Nous nous intéressons dans un premier

temps, dans ce travail, à l'extraction des intérêts de l'utilisateur (et non à leur mise à jour) dans le but d'obtenir un profil social pertinent dès sa première construction.

4. Proposition

Notre travail se fonde sur des critères temporels pour prendre en compte la dynamique de la structure du réseau mais aussi la dynamique des informations échangées dans le processus de construction du profil social de l'utilisateur.

Dans notre approche, l'unité d'intérêt de base correspond à un élément (terme d'un titre, terme d'un document, etc.). Ces éléments sont en général le résultat d'un traitement de l'information brute utilisée par élimination des mots vides, lemmatisation, synonymie, etc.

Nous proposons d'intégrer un poids temporel dans l'étape de pondération des intérêts dans le processus (décrit dans la section suivante) de construction du profil social. La pondération temporelle des intérêts consiste à attribuer aux éléments extraits, un poids prenant en compte des critères temporels (poids temporel) que l'on appelle « TempScore » (pour *Score Temporel*).

Nous donnons ci-dessous l'algorithme de construction du profil social appelé CoBSPT (*Community Based Social Profile - T* pour « Temporel ») qui est basé sur le processus CoBSP (présenté section 2.1) en intégrant le calcul du poids temporel. Par rapport à la méthode CoBSP, ce travail se focalise sur l'intégration d'une information temporelle dans l'étape de profilage de communautés (étape 2 de l'algorithme CoBSP), et dans l'étape de caractérisation sémantique (étape 3).

Dans cet algorithme, nous avons en entrée le réseau égo-centrique de l'utilisateur duquel on extrait les communautés. Ensuite, le profil de chaque communauté est calculé. Tout d'abord, l'algorithme extrait les éléments (termes) à partir des informations partagées par tous les individus présents dans la communauté. Nous appliquons à cette étape des techniques classiques d'élimination de mots vides et de recherche de synonymes. Nous calculons le poids temporel de chaque élément pour chaque individu. Le poids temporel est calculé en combinant le poids temporel des individus (fonction *calculIndivTempScore*) et le poids temporel des informations (fonction *calculInfoTempScore*). Ces poids au niveau des individus sont agrégés au niveau de la communauté. Chaque élément de chaque communauté est alors défini par son poids sémantique et structurel (comme défini dans l'algorithme CoBSP). Enfin, on agrège les éléments de toutes les communautés pour obtenir l'ensemble des éléments dans le profil social qui représentent les intérêts de l'utilisateur sous forme d'un vecteur de termes pondérés.

Nous considérons cet algorithme comme générique car certaines fonctions peuvent être personnalisées selon les techniques et le jeu de données (RSN) choisis comme nous le verrons par la suite.

Dans l'étape de profilage d'une communauté, nous proposons de calculer le poids temporel de chaque élément/terme, d'une part, à partir de la pertinence temporelle « *InfoTempScore* » des informations utilisées pour extraire cet élément et, d'autre part, à partir de la pertinence « *IndivTempScore* » de l'individu qui partage ces informations. La pertinence temporelle de l'information est calculée en prenant en compte la fraîcheur de l'information. La pertinence de l'individu est calculée en utilisant la force des relations entre l'individu en question et l'utilisateur central. Nous combinons ensuite ces deux poids pour obtenir le poids temporel de chaque élément. L'algorithme est détaillé ci-après et les différentes fonctions de calculs sont expliquées dans les sections suivantes.

Algorithme CoBSPT– Construction du profil social à partir du réseau égocentrique basé sur les communautés en prenant en compte les critères temporels

Notations utilisées :

- k dénote le réseau égocentrique de l'utilisateur
- C_i dénote une communauté du réseau égocentrique.
- $C = \{C_i\}$ est l'ensemble des communautés du réseau égocentrique.
- Un élément e dénote un élément (terme) extrait d'une information (après traitement classique d'élimination de mots vides, recherche de synonymes, lemmatisation, ...).
- $E(C_i)$ dénote l'ensemble des éléments (termes) extraits des informations de tous les individus d'une communauté C_i .
- Les notations P_{Temp}^{xxx} dénotent des calculs de poids temporels (Temp) utilisés dans l'algorithme et sont détaillés dans les sections suivantes.
- *combinaison* (a, b, δ) dénote une fonction de combinaison de la forme : $a.\delta + (1-\delta).b$.

1 : Algorithme CoBSPT

2 : Début

3 : //1ère étape : détection de communautés dans le réseau égocentrique k

4 : $C := iLCD(k)$

5 : //2ème étape : profilage des communautés

6 : Pour chaque communauté C_i appartenant à C , Faire

7 : Pour chaque individu $Indiv$ de C_i Faire

8 : $P_{Temp}^{Indiv} = \text{calculIndivTempScore}(Indiv)$;

9 : Pour chaque information $Info$ partagée par $Indiv$ Faire

10 : $P_{Temp}^{Info, Indiv} = \text{calculInfoTempScore}(Info, Indiv)$;

11 : $P_{Temp}^{Indiv, Info}(e) = \text{combinaison}(P_{Temp}^{Info, Indiv}, P_{Temp}^{Indiv, \gamma})$ | e est un élément $\in Info$

12 : Fin pour

13 : Fin pour

14 : // Calcul du poids de chaque e dans la communauté C_i par agrégation des

15 : // $P_{Temp}^{Indiv, Info}(e)$ dans l'ensemble des éléments du profil de la communauté

16 : Pour chaque e appartenant à $E(C_i)$, Faire

17 : // pour toute $Info$ de tous les individus $Indiv$ de la communauté C_i ;

18 : $P_{Temp}^{C_i}(e) = \text{PoidsAgrégation}(P_{Temp}^{Indiv, Info}(e), C_i)$

19 : Fin pour

20 : Fin pour

21 : //3° étape : caractérisation sémantique et structurelle des communautés

22 : Pour chaque communauté C_i appartenant à C , Faire

23 : $PStruct(C_i) = \text{calculMesureCentralité}(C_i)$;

```

24 :      Pour chaque e appartenant à E(Ci), Faire
25 :          PSem(e, Ci) =  $\frac{P_{Temp}^{C_i}(e)}{\sum_{C_j \in C} P_{Temp}^{C_j}(e)}$ 
26 :          PSemStruct(e, Ci) = combinaison (PStruct(Ci), PSem(e, Ci),  $\alpha$ )
27 :      Fin pour
28 :  Fin pour
29 :  //4e étape : dérivation du profil social
30 :  Pour chaque élément e appartenant à  $\cup E(C_i) \forall C_i \in C$  Faire
31 :      PSocial(e) = combinLin(PSemStruct(e, Ci)) ;
32 :  Fin pour
33 :  Retourner les couples (e, PSocial(e)) pour tout e appartenant à  $\cup E(C_i) \forall C_i \in C$ ;
34 : Fin

```

5. Calcul du poids temporel *TempScore*

Nous présentons dans cette section, le calcul du poids temporel des intérêts « *TempScore* » qui est calculé en combinant le poids de pertinence temporel des individus et des informations. Nous détaillons dans la suite, le calcul de ces deux poids avec les techniques choisies ainsi que la combinaison de ces deux poids pour obtenir le poids final.

5.1. Calcul du poids temporel des informations (*InfoTempScore*)

5.1.1. Analyse des fonctions temporelles existantes

Pour pondérer des informations selon des critères temporels, plusieurs travaux utilisent les fonctions temporelles. La fonction la plus utilisée est la fonction exponentielle. (Ding et Li, 2005) proposent une fonction exponentielle (2) pour pondérer les informations afin de les utiliser dans un système de recommandation temporelle.

$$f(t) = e^{-\lambda t} \quad (2)$$

La valeur t représente la fraîcheur de l'information. Pour chaque $t=i$ ($i \in \mathbb{N}$), $t=0$ est considéré comme la valeur de fraîcheur de l'instant le plus récent (ex : $t=0$ pour l'année 2014, $t=1$ pour 2013, etc.). La valeur $\lambda \in [0,1]$ représente le taux de décroissance (*Time Decay Rate*) des valeurs. Plus λ est grand, moins les informations anciennes sont importantes. Nous présentons dans la section 5 une étude paramétrique complète nous ayant permis de sélectionner une valeur pour λ .

5.1.2. Calcul du poids temporel des informations

Nous décrivons dans cette section la fonction calculInfoTempScore de l'algorithme CoBSPT. A partir des informations *Info* partagées par les individus du réseau égocentrique de l'utilisateur, nous proposons de pondérer leur poids de pertinence en se basant sur la fonction temporelle $f(t)$ présentée dans l'équation (2). Dans notre contexte, nous proposons d'appliquer différemment la valeur fraîcheur t

dans cette formule. En effet, en se basant sur la caractéristique dynamique du réseau social, nous supposons que les informations partagées au moment de la dernière interaction entre l'individu et l'utilisateur principal u sont les informations les plus pertinentes pour l'utilisateur u . De ce fait, l'importance des informations partagées avant et après une interaction sera diminuée selon la distance temporelle entre la date de partage de ces informations et la date de la dernière interaction. La fraîcheur t pour la fonction (2) sera donc calculée par rapport à la date de dernière interaction entre l'individu $Indiv$ et l'utilisateur central u . Dans ce cas, $t=0$ représente l'estampille de publication la plus fraîche par rapport à la date de la dernière interaction entre l'individu $Indiv$ et l'utilisateur central u . Par exemple si la date de la dernière interaction entre u et $Indiv$ est en 2015, la valeur t pour les informations partagées en 2015 = 0 et la valeur t pour les informations partagées en 2014 et 2013 = 1. La valeur de t est calculée par la fonction (3).

$$t_{Info} = |(date\ dernière\ interaction\ (Indiv,u)) - (date\ de\ publication\ de\ Info)| \quad (3)$$

Le poids temporel d'un terme e extrait depuis une information $Info$ est calculé avec la formule (4) ci-dessous :

$$P_{Temp}^{Info,Indiv}(e) = f(t_{Info}) \quad (4)$$

5.2. Calcul du poids temporel des relations (*IndivTempScore*)

5.2.1. Techniques existantes de calcul du poids de pertinence des liens

Il existe plusieurs mesures pour calculer le poids de pertinence entre deux individus dans le réseau social. Parmi elles, on peut citer la fréquence d'interactions entre deux utilisateurs. Par exemple, dans le domaine de co-publications scientifiques, on peut utiliser la fréquence de co-publications. On peut également exploiter cette mesure pour calculer l'exclusivité par publication qui donne plus de poids aux relations entre co-auteurs de publications ayant un nombre faible de co-auteurs par rapport aux publications ayant un grand nombre de co-auteurs (Cabanac, 2011). Dans les RSN, pour calculer la force entre deux nœuds, les mesures de similarité qui se basent sur la topologie du réseau sont utilisées. Nous pouvons trouver ces mesures dans les travaux de prédiction de liens (Liben-Nowell et Kleinberg, 2003). Nous proposons dans notre travail d'appliquer les mesures utilisées dans le domaine de la prédiction de liens pour calculer un poids de similarité entre deux nœuds déjà connectés qui représentera la persistance du lien entre ces nœuds dans le futur.

Nous nous intéressons en particulier dans ce travail à la mesure de prédiction de liens *Adamic/Adar*, introduite dans (Liben-Nowell et Kleinberg, 2003) qui s'est avérée performante malgré sa simplicité. Cette technique se base sur le nombre de voisins communs en donnant des poids plus importants aux voisins communs les

plus rares. Pour une paire de nœuds (x,y) , la formule de calcul de score de similarité de *Adamic/Adar* est :

$$\text{AdamicAdar}(x,y) = \sum_{z \in \{\Gamma(x) \cap \Gamma(y)\}} \frac{1}{\log|\Gamma(z)|} \quad (5)$$

où $\Gamma(x)$ est l'ensemble des voisins de x .

La plupart des travaux s'intéressent à la prédiction de liens sans tenir compte de critères temporels. Cependant, comme introduit précédemment, la prise en compte du temps est essentielle pour un RSN. Nous nous intéressons donc à la prédiction de liens temporelle qui prend en compte les informations structurelles et temporelles lors du calcul du score de similarité entre deux nœuds. (Tylenda *et al.*, 2009) proposent une technique de prédiction temporelle qui applique des facteurs temporels dans le calcul des poids de pertinence des individus. En se basant sur ce travail, la formule de *Adamic/Adar* qui prend en compte les informations temporelles devient :

$$\text{AdamicAdarTemps}(x,y) = \sum_{z \in \{\Gamma(x) \cap \Gamma(y)\}} \frac{w(x,z) \cdot w(z,y)}{\log|\Gamma(z)|} \quad (6)$$

où $\Gamma(x)$ est l'ensemble des voisins de x . La fonction $w(a,b)$ dans la formule (6) représente le poids temporel entre deux nœuds donnés a et b . Cette fonction est personnalisable selon les techniques et le jeu de données choisis. Nous détaillons cette fonction dans la partie suivante selon les techniques choisies.

5.2.2. Calcul de poids d'individus du réseau égocentrique (*IndivTempScore*)

Nous décrivons dans cette section la fonction `calculIndivTempScore` de l'algorithme `CoBSPT`. Pour calculer le poids temporel des individus, nous considérons que les individus qui ont les relations les plus récentes avec l'utilisateur central ont une probabilité plus grande de partager les mêmes intérêts avec cet utilisateur (plus significatifs). Nous donnons donc un poids plus important aux informations partagées par ces individus par rapport aux individus ayant des liens moins récents. Il s'agit de calculer la pertinence d'un individu *Indiv* par rapport à l'utilisateur central u en prenant en compte les informations temporelles de ses liens avec l'utilisateur central (par exemple, la date de création de liens avec l'utilisateur central, la durée de la relation, etc.).

Nous proposons d'appliquer la fonction temporelle $f(t)$ présentée dans la formule (2) aux relations entre chaque individu *Indiv* et l'utilisateur central u . Nous choisissons donc $w(u, \text{Indiv}) = f(\text{dateMaintenant} - \text{interaction}(u, \text{Indiv}))$ où $\text{interaction}(u, \text{Indiv})$ représente la date de la dernière interaction entre u et *Indiv*. Le poids temporel entre l'utilisateur principal u et l'individu *Indiv* est calculé en appliquant la formule (6). Dans ce contexte t représente la fraîcheur de la date de publication ($t=0$ étant la plus récente). Le poids *temporel* d'un terme e extrait à partir

d'une information partagée par l'individu *Indiv* est donc calculé par la formule (7) ci-dessous :

$$P_{Temp}^{Indiv} = AdamicAdarTemps(u, Indiv) \quad (7)$$

Notons que ce poids est identique pour tous les éléments *e* appartenant à chaque *Info* de l'individu *Indiv*.

5.3. Poids final (TempScore) et dérivation du profil social

Pour obtenir le poids final d'une information, les poids temporels des informations et des individus sont combinés. Le poids d'un élément *e* extrait depuis un individu *Indiv* du réseau égocentrique d'un utilisateur central sera calculé via la formule suivante :

$$P_{Temp}^{Indiv,Info}(e) = \gamma \cdot (P_{Temp}^{Indiv}) + (1-\gamma) \cdot (P_{Temp}^{Info,Indiv}) \mid e \in Info \quad (8)$$

Le paramètre γ (compris entre 0 et 1) dans la formule permettra de jauger et de faire varier l'importance du poids temporel de l'individu dans la dérivation du profil social de l'utilisateur par rapport au poids de l'information. Plus la valeur de γ est importante plus on donne de l'importance à ce poids. Nous présentons dans la section 5 une étude paramétrique complète nous ayant permis de sélectionner une valeur satisfaisante pour γ .

Une fois que le poids temporel de chaque élément ($P_{Temp}^{Indiv,Info}(e)$) est calculé. Nous agrégeons les éléments trouvés dans chaque communauté C_i (somme des poids de tous les termes identiques) en calculant leur poids par rapport au poids total de tous les termes (fonction PoidsAgrégation de l'algorithme CoBSP, calculé avec la formule (9)).

$$PoidsAgrégation(e, C_i) = \frac{\sum P_{Temp}^{Indiv,Info}(e)}{\sum_{f \in E(C_i)} P_{Temp}^{Indiv,Info}(f)} \quad (9)$$

$E(C_i)$ dénote l'ensemble des éléments (termes) extraits des informations de tous les individus d'une communauté C_i .

Dans l'étape 3, l'algorithme CoBSP est modifié par l'intégration du poids temporel calculé dans l'étape 2 pour la caractérisation sémantique des communautés. Nous appliquons ensuite les mêmes méthodes que CoBSP pour dériver le profil. Nous obtenons finalement un profil social représenté sous forme d'un vecteur des termes pondérés.

6. Expérimentations

Pour valider notre proposition, nous avons comparé la pertinence de notre approche par rapport à l'approche existante. Pour cela, nous comparons la

pertinence des profils sociaux construits par ces différentes approches. Le domaine d'expérimentations choisi concerne les réseaux d'auteurs de publications scientifiques (DBLP ⁵ et Mendeley⁶). Dans ces réseaux, les nœuds représentent les auteurs. Deux nœuds peuvent être reliés par un lien s'ils publient ensemble. Nous exploitons le réseau des auteurs DBLP pour lesquels on peut calculer les centres d'intérêt à partir des titres de leurs publications, on peut alors construire le réseau égocentrique de chaque auteur et dériver son profil social. Pour calculer la pertinence de ce profil social construit, nous le comparons avec son profil explicite (profil utilisateur) qui contient les intérêts réels de l'utilisateur que nous dérivons à partir des intérêts indiqués explicitement par l'utilisateur dans son profil Mendeley.

Nous partons des travaux existants (CoBSP) pour construire les profils sociaux et extraire le profil explicite de l'utilisateur. Comme présenté dans la figure 2, nous recherchons dans un premier temps, les auteurs de publications scientifiques dans Mendeley ayant plus de 6 centres d'intérêt pour avoir un profil explicite et significatif de l'utilisateur. Ensuite, nous partons de ces auteurs pour extraire leur réseau égocentrique depuis leurs données dans DBLP. Notre échantillon de test porte sur 115 auteurs ayant entre 27 et 209 co-auteurs.

Nous construisons ensuite pour un auteur, les différents profils sociaux à partir de son réseau égocentrique DBLP (issus des différentes approches de construction de profil social). D'une part, le profil social est calculé selon la méthode existante CoBSP. D'autre part, le profil social est calculé selon notre approche. Nous évaluons ensuite les deux profils sociaux en les comparant au profil explicite extrait de Mendeley. Nous utilisons les mesures de rappel et de précision classiques dans lesquelles les intérêts des profils remplacent les documents recherchés ou trouvés. Nous considérons les intérêts du profil explicite Mendeley intérêts pertinents (versus documents pertinents) et les intérêts des profils sociaux construits comme des intérêts retrouvés (versus documents retrouvés).

6.1. Construction du profil social

Dans cette étape, nous construisons les profils sociaux avec les différents algorithmes (algorithme CoBSP pour l'approche existante et algorithme CoBSPT pour l'approche proposée).

5. dblp.uni-trier.de

6. www.mendeley.com

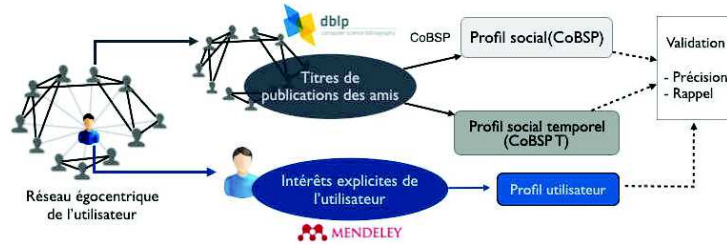


Figure 2. Illustration de la construction du profil social et du profil explicite de l'utilisateur depuis des données DBLP et Mendeley

6.1.1. Construction du profil social pour l'algorithme existant (CoBSP)

Comme décrit précédemment, le réseau égocentrique d'un auteur (ego) est construit à partir des relations entre les co-auteurs fournies par DBLP. La première étape consiste à récupérer les co-auteurs de cet ego et construire les communautés. Nous stockons ensuite la liste des publications de tous les auteurs qui font partie de son réseau égocentrique. Ceci permet de récupérer les titres des publications des communautés. Nous analysons dans la deuxième étape, les titres des publications pour en extraire les termes les plus significatifs qui seront considérés comme les centres d'intérêt de l'utilisateur. Nous utilisons un dictionnaire de synonymes pour chaque terme. Ces synonymes sont ensuite considérés comme des occurrences du terme concerné. Les filtres permettent de sélectionner les mots les plus importants et d'en exclure les moins significatifs (tels que des mots vides). Seuls les mots retenus seront considérés comme des centres d'intérêt de l'utilisateur. Enfin, dans la troisième étape, nous exploitons les mesures sémantiques et les mesures structurelles et représentons les centres d'intérêt de ce profil social par un vecteur de termes pondérés (Tchuente *et al.*, 2013).

6.1.2. Construction du profil social en prenant en compte des critères temporels (CoBSPT)

Pour construire le profil social à partir de notre approche, nous utilisons le même principe que celui du processus CoBSP mais en appliquant les critères temporels selon la technique de calcul de poids temporel proposée. Pour chaque terme trouvé dans chaque communauté, nous calculons le poids temporel *TempScore* par la formule (8). Comme défini dans la section 5, le profil social construit est appelé CoBSPT.

6.2. Extraction du profil explicite de l'utilisateur à partir de Mendeley

Pour construire le profil utilisateur, nous utilisons les intérêts indiqués explicitement par l'utilisateur dans son profil Mendeley. En appliquant le même

traitement que celui de la construction du profil social, nous utilisons les dictionnaires et les filtres pour extraire des termes significatifs qui seront considérés comme les centres d'intérêt dans le profil explicite de l'utilisateur. A la différence du profil social, la pondération de chaque mot est définie uniquement par la mesure de fréquence tf dans le texte entier analysé. Les centres d'intérêt de ce profil seront également représentés par un vecteur de termes pondérés.

6.3. Evaluation

Après l'étape de construction des profils sociaux et du profil utilisateur, nous obtenons 2 profils sociaux différents construits par les algorithmes CoBSP et CoBSPT. Les profils sociaux construits seront comparés au profil explicite de l'utilisateur via les mesures de précision et de rappel. Dans notre contexte d'évaluation, la précision d'un algorithme de dérivation du profil social est évaluée par le nombre de centres d'intérêt prédits (profil social) trouvés dans le profil utilisateur par rapport au nombre total des centres d'intérêt calculés dans le profil social.

$$Précision = \frac{nb(intérêts\ dans\ le\ profil\ social \cap intérêts\ dans\ le\ profil\ utilisateur)}{nb(intérêts\ dans\ le\ profil\ social)} \quad (10)$$

Le rappel d'un algorithme de dérivation du profil social quant à lui est évalué par le nombre de centres d'intérêt prédits (profil social) qui sont trouvés dans le profil utilisateur par rapport au nombre de centres d'intérêt du profil utilisateur.

$$Rappel = \frac{nb(intérêts\ dans\ le\ profil\ social \cap intérêts\ dans\ le\ profil\ utilisateur)}{nb(intérêts\ dans\ le\ profil\ utilisateur)} \quad (11)$$

Pour calculer la précision et le rappel, nous nous intéressons uniquement aux centres d'intérêt les plus pertinents renvoyés par chaque algorithme de dérivation du profil social. Nous retenons ainsi les top n premiers centres d'intérêt, donc un vecteur de taille n pour représenter le profil social. En effet, les profils sociaux construits sont de grande taille (en moyenne 2031 éléments) et ne seraient pas exploitables en l'état

6.4. Résultats

Nous avons lancé plusieurs expérimentations avec différentes combinaisons de valeurs de γ , λ et $\alpha \in \{0.01, 0.05, 0.1, 0.25, 0.5, 0.75, 0.9, 0.95, 1.0\}$ pour étudier l'impact de la dynamique d'informations dans le réseau par rapport à celle de la structure du réseau.

Dans un premier temps, nous calculons les résultats des profils avec la valeur $\gamma = 1$ qui représente le profil calculé en prenant en compte seulement le poids temporel des individus (*IndivTempScore*) et qui représente la prise en compte de la

dynamique de la structure du réseau. La valeur de λ varie pour trouver la valeur qui donne les meilleurs résultats. Selon les résultats présentés dans le tableau 1, en calculant la moyenne de précision et de rappel pour les top 10 des intérêts pour toutes les valeurs de α , nous trouvons le meilleur résultat quand $\lambda = 0,75$ qui donne respectivement un gain de 11,2 % et 7,5 % en termes de précision et de rappel par rapport à l'algorithme CoBSP (qui donne 0,1803 et 0,1738 en termes de précision et de rappel). Cela nous montre dans un premier temps, l'intérêt d'utiliser une méthode de prédiction de liens temporelle pour sélectionner les individus pertinents pour le processus de construction du profil social.

Nous avons ensuite calculé les résultats avec la valeur $\gamma = 0$ qui représente le profil calculé en prenant en compte seulement le poids temporel des informations (*InfoTempScore*). Nous avons fait varier également la valeur de λ pour trouver la valeur qui donne les meilleurs résultats. Nous obtenons le meilleur résultat par rapport à CoBSP (qui donne 0,1798 et 0,17 en termes de précision et de rappel) quand $\lambda = 0,01$ (voir le tableau 1) avec un gain de 1,2 % et 7 % en termes de précision et de rappel pour les top 10 des intérêts. Cela nous montre l'importance de prendre en compte les critères temporels pour sélectionner des informations et construire le profil social.

*Tableau 1. Comparatif des mesures de rappel et de précision de différentes valeurs de λ pour les calculs du poids temporel d'individus (*IndivTempScore*) ($\gamma = 1$) et du poids temporel d'informations (*InfoTempScore*) ($\gamma = 0$). Les valeurs présentées sont des moyennes sur tous les paramètres α*

λ	IndivTempScore($\gamma=1$)		InfoTempScore ($\gamma=0$)	
	Précision	Rappel	Précision	Rappel
0	0.16546583850931	0.1626480933969	0.1776535541753	0.1776535541753
0.01	0.16621118012422	0.1617124043008	0.1819737750173	0.1819737750172
0.05	0.16571428571429	0.1634325080715	0.1669427191166	0.1669427191166
0.1	0.16968944099379	0.1640187441627	0.1524361628709	0.1524361628710
0.2	0.18111801242236	0.1716781957609	0.1416425120773	0.1416425120773
0.5	0.19180124223602	0.1811317032275	0.1243478260869	0.1243478260870
0.75	0.20049689440994	0.1868252809135	0.1169358178054	0.1169358178054
0.9	0.19900621118012	0.1851941277010	0.1146583850932	0.1146583850932
1	0.20024844720497	0.1843087417820	0.1134851621808	0.1134851621808

Nous avons ensuite calculé les résultats finaux avec la combinaison de deux poids en fixant la valeur de $\lambda = 0,75$ pour le calcul du poids temporel des individus (*IndivTempScore*) et $\lambda = 0,01$ pour le poids temporel des informations

(InfoTempScore). Nous avons fait varier la valeur de γ pour trouver la meilleure combinaison en se basant sur la valeur moyenne pour tous les α . Selon les résultats de le tableau 2, la valeur optimale de γ est à 0,75. Avec ce paramétrage nous trouvons de meilleurs résultats par rapport à ceux de l'approche CoBSP en termes de précision et de rappel.

Tableau 2. Comparatif des mesures de rappel et de précision avec différentes valeurs de γ en fixant la valeur de $\lambda = 0,75$ pour le poids temporel des individus (IndivTempScore) et $\lambda = 0,01$ pour le poids temporel des informations (InfoTempScore)

γ	Précision	Rappel
0	0.18100759144237155	0.18630755508353936
0.1	0.18106280193236454	0.18625356358727535
0.2	0.18048309178743707	0.1853751872197925
0.25	0.18162870945479384	0.1864698830623874
0.5	0.18222222222221954	0.18777071139006993
0.75	0.18461007591442075	0.19031104571895474
0.9	0.18251207729468283	0.18811154825947243
1	0.18245686680468987	0.18552500652619888

La figure 3 représente les résultats pour les top 5 et top 10 des intérêts permettant de comparer le profil social construit avec ce dernier paramétrage et le profil social de l'approche existante. Dans cette figure, la valeur α représente la proportion entre la mesure sémantique et la mesure structurelle de chaque communauté comme présentée dans la formule (1).

Par rapport à la figure 3, l'algorithme CoBSPTP produit de meilleurs résultats que l'algorithme CoBSP en termes de précision et de rappel pour toutes les valeurs de α , en particulier lorsque CoBSP obtient ses meilleurs résultats. Pour le top 5 des centres d'intérêts, le meilleur résultat peut être observé quand $\alpha = 0.0$ avec successivement 16,6 % et 16,8 % de taux de gain en termes de précision et rappel par rapport à l'algorithme CoBSP (qui donne 0,1878 et 0,1 en terme de précision et de rappel). Pour le top 10 des centres d'intérêts, le taux de gain en termes de précision et rappel est respectivement à 14,06 % et 16,7 % quand le meilleur résultat est observé avec $\alpha = 0,0$ par rapport à l'algorithme CoBSP qui donne 0,1669 et 0,178 en termes de précision et de rappel. Il semble alors que la prise en compte de l'aspect structurel lors de l'agrégation des profils de communautés soit sans apport d'information nouvelle à l'opposé de la prise en compte de la dynamique du réseau mais au niveau de chaque communauté (avant agrégation).

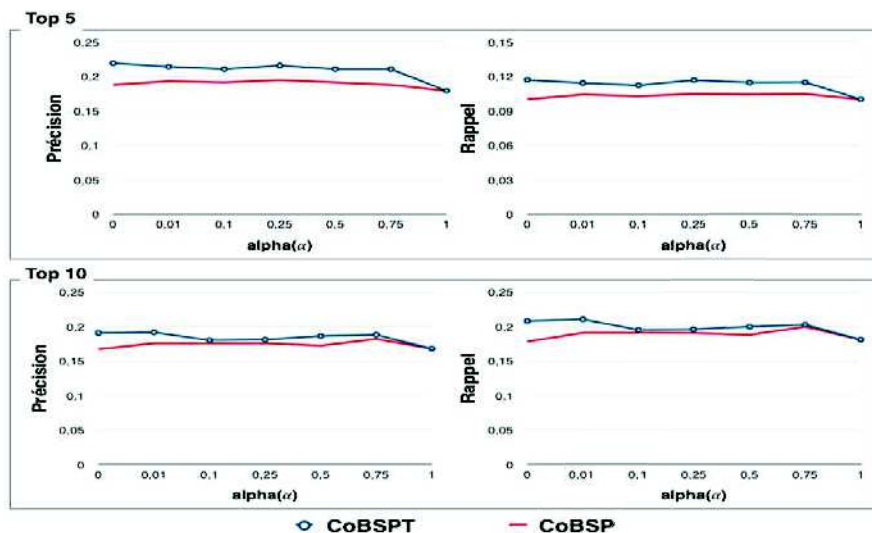


Figure 3. Comparatif de la pertinence des profils sociaux construits avec CoBSP et CoBSPT par rapport au profil explicite de l'utilisateur. $\gamma = 0,75$, $\lambda = 0,75$ pour le poids temporel des individus (IndivTempScore) et $\lambda = 0,01$ pour le poids temporel des informations (InfoTempScore)

Nous avons ensuite découpé l'échantillon en plusieurs intervalles pour pouvoir mieux observer le résultat par rapport au nombre de co-auteurs des auteurs étudiés. Nous analysons les 3 sous-échantillons suivants :

- 30 auteurs ayant moins de 50 co-auteurs représentant les auteurs d'un réseau *épars* soit 26,08 % de l'échantillon,
- 60 auteurs ayant entre 50 et 100 co-auteurs représentant les auteurs d'un réseau *assez dense* soit 52,17 % de l'échantillon,
- 25 auteurs ayant plus de 100 co-auteurs représentant les auteurs d'un réseau *dense* soit 21,73 % de l'échantillon.

La figure 4 présente le comparatif de la pertinence des profils sociaux pour chaque intervalle du nombre de co-auteurs avec le top 5 des centres d'intérêts.

Pour les auteurs possédant moins de 50 co-auteurs :

- Si on s'intéresse aux deux algorithmes : le meilleur écart de résultat s'obtient lorsque α vaut 0,75. Cela donne un gain de précision de 58 % et un gain de rappel de 28 %. Le gain moyen pour toutes les valeurs de α est de 28 % pour la précision et de 29 % pour le rappel.
- Si on se concentre sur CoBSPT : les résultats les meilleurs sont obtenus pour α égale 0,0 ce qui rejoint le premier résultat global précédent.

Pour les auteurs possédant entre 50 et 100 co-auteurs :

– Si on s'intéresse aux deux algorithmes : le meilleur écart de résultats s'obtient ici encore lorsque α vaut 0,75. Cela donne un gain de précision de 20 % et un gain de rappel de 20 %. Le gain moyen pour toutes les valeurs de α est de 12 % pour la précision et de 11 % pour le rappel.

– Si on se concentre sur CoBSPT : là encore les résultats les meilleurs sont obtenus pour α égale 0,0 ce qui rejoint le premier résultat global précédent.

Pour les auteurs possédant plus de 100 co-auteurs :

– Si on s'intéresse aux deux algorithmes : le meilleur écart de résultats s'obtient lorsque α vaut 0,0. Cela donne un gain de précision de 11 % et un gain de rappel de 12 %. Néanmoins, en moyenne, CoBSPT est moins performant : la perte moyenne pour toutes les valeurs de α est de 7 % pour la précision et de 3 % pour le rappel.

– Si on se concentre sur CoBSPT : le premier résultat global de meilleure performance de l'algorithme pour α égale à 0,0 est confirmé et les performances sont très comparables (mais inférieures) aux meilleurs résultats de CoBSP ($\alpha = 0,25$). Ce résultat est cohérent compte tenu des observations de (Tchunte *et al.*, 2013) donnant de très bons résultats sur les réseaux denses avec α autour de 0,25.

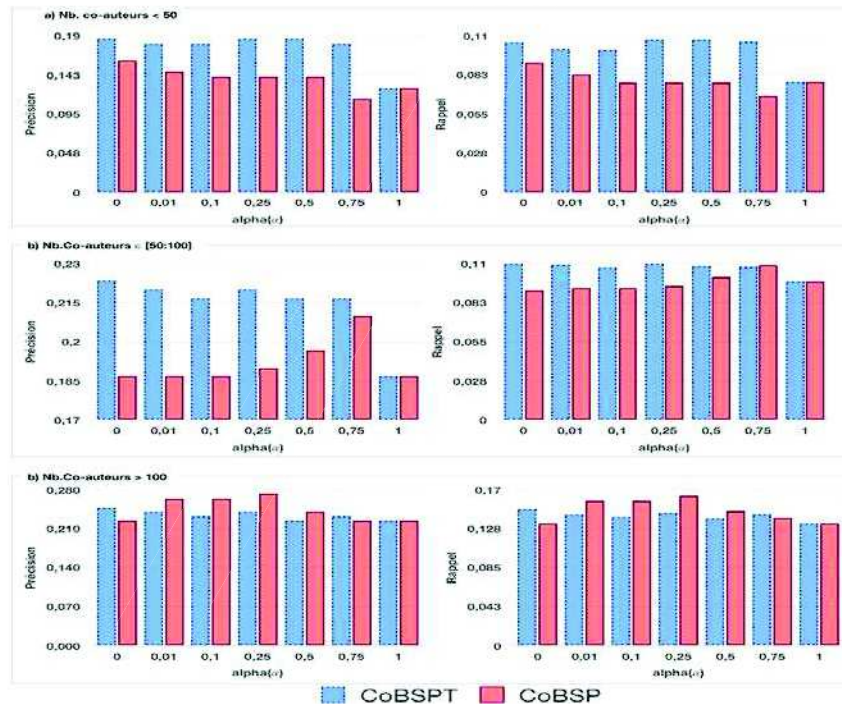


Figure 4. Comparatif des mesures de précision et de rappel des profils sociaux construits avec CoBSP, CoBSPT pour différents intervalles du nombre de co-auteurs

6.5. Discussion de résultats

Nos expérimentations montrent l'efficacité de notre approche par rapport à l'approche existante et montre par conséquent l'importance de la prise en compte de critères temporels non seulement dans la sélection des sources d'informations liées à la structure du réseau social de l'utilisateur mais aussi dans le traitement des informations partagées dans le réseau. Cependant, on remarque que, pour obtenir de meilleurs résultats, la valeur de λ (qui représente le taux de décroissance) utilisée pour calculer le poids temporel des informations doit être plus faible que celle utilisée pour calculer le poids temporel des individus.

Nous pouvons expliquer ces résultats par rapport au type du réseau social étudié dans ces expérimentations : le réseau de publications scientifiques qui ne représente pas forcément un réseau avec une évolution rapide des informations partagées. Comme nous utilisons les titres de publications comme informations partagées dans le réseau, il se peut que le taux de changement de domaine de publications scientifiques des auteurs soit assez bas par rapport aux changements d'informations partagées dans d'autres types de réseaux. Par exemple, un utilisateur peut mettre un an à publier un article alors que dans un réseau de partage d'informations sociales comme Twitter, les informations peuvent être partagées toutes les minutes. De plus, on ne rencontre pas dans le réseau de publications scientifiques DBLP une diversité d'informations qui pourraient évoluer ou changer à travers le temps comme celles des réseaux de partage d'informations. Un auteur peut mettre entre 5 et 10 ans pour changer de domaine de recherche tout en restant dans le domaine de l'informatique alors que dans les réseaux de partage d'informations sociales, les utilisateurs peuvent partager rapidement les informations sur différents sujets.

Nous envisageons donc dans nos prochaines expérimentations d'appliquer notre approche sur d'autres types de réseaux sociaux où la caractéristique dynamique des informations est très importante (Facebook, Twitter, etc.) afin d'étudier son impact sur ces réseaux et ainsi renforcer notre proposition dans d'autres contextes de RSN.

Par rapport à l'analyse par nombre de co-auteurs, nous constatons que notre approche peut améliorer, avec un gain important, la pertinence du profil social dans le cas de réseaux plutôt épars. Cependant dans les réseaux qui sont très denses, notre approche n'améliore les résultats que quand α vaut 0.0 et avec un gain moins important par rapport à ceux d'autres sous-échantillons. Nous pouvons expliquer ces résultats d'abord, par le fait que les réseaux peu denses contiennent moins de relations dans le réseau et fournissent moins d'informations significatives sur les liens et même sur les informations partagées. Ce problème a également été soulevé dans (Tchunte *et al.*, 2013) qui trouve de bons résultats pour le processus CoBSP quand le réseau est assez dense (utilisateurs ayant de 50 à 100 co-auteurs ou plus de 100 co-auteurs). En appliquant la technique de prédiction de liens temporelle, nous pouvons sélectionner les liens de manière plus significative et donc améliorer ces résultats.

Un facteur qui pourrait changer les résultats pour chaque sous-échantillon peut être lié aux valeurs des paramètres λ et γ que l'on pourrait faire varier selon la taille et le type du réseau. Dans cette expérimentation, nous utilisons les mêmes valeurs de λ et γ pour les 3 sous-échantillons. Nous envisageons de faire varier les valeurs de ces paramètres selon le type et les caractéristiques du réseau dans de futures expérimentations.

7. Conclusion et perspectives

Dans ce travail, nous proposons de prendre en compte la caractéristique dynamique des réseaux sociaux dans l'étape d'extraction des intérêts du processus de construction du profil social. Dans ce processus, nous intégrons des critères temporels pour pondérer les individus et les informations (poids temporel) du réseau social de l'utilisateur afin de pouvoir extraire des intérêts de l'utilisateur pertinents et à jour. Ce poids temporel est calculé, à partir du poids de pertinence temporel des individus et du poids de pertinence temporel des informations que ces individus partagent. Le poids de pertinence temporel des individus est calculé en appliquant une méthode de prédiction de liens temporelle afin de sélectionner les individus ayant les liens les plus actifs avec l'utilisateur. Le poids de pertinence temporel des informations partagées est calculé en prenant en compte la fraîcheur des informations. Les résultats des expérimentations nous ont permis de montrer l'efficacité de cette approche par rapport au processus de construction du profil social qui ne tient pas compte de critères temporels.

À court terme, nous envisageons d'appliquer cette approche dans d'autres types de RSN (Facebook, Twitter, etc.) pour effectuer une évaluation de notre proposition à plus grande échelle. Nous allons également étudier notre approche selon différentes caractéristiques du réseau (par ex. taille et densité). Nous souhaitons également étudier d'autres algorithmes de prédiction de liens temporelle ainsi que d'autres fonctions temporelles afin de tenter d'améliorer la performance de notre approche. À plus long terme, nous envisageons la mise à jour continue du profil social, afin d'avoir un profil pertinent et à jour à tout moment, pour des mécanismes d'adaptation de l'information par exemple.

Un autre phénomène important qui peut avoir un impact sur la performance du profil social est l'apparition et la propagation de rumeurs dans le réseau social. En effet, dans (Mezghani *et al.*, 2015), nous avons montré comment détecter pour anticiper la propagation de *buzz* ou rumeurs dans un système de recommandation. En effet, ne pas distinguer les rumeurs des vraies informations amène à les exploiter pour extraire les intérêts et peut conduire à proposer de faux intérêts dans le profil, et par conséquent donner de fausses interprétations pour les mécanismes d'adaptation de l'information. Il s'agira alors de mettre en œuvre des mécanismes des détections et de filtrage de ces informations.

Bibliographie

- Abel F., Gao Q., Houben G.-J., Tao K. (2011). Analyzing Temporal Dynamics in Twitter Profiles for Personalized Recommendations in the Social Web, *Proceedings of the 3rd International Web Science Conference, WebSci'11*. ACM, New York, NY, USA, p. 2:1-2:8. doi:10.1145/2527031.2527040
- Aggarwal C., Subbian K., (2014). Evolutionary Network Analysis: A Survey. *ACM Comput Surv* 47, vol. 47, n°1, doi:10.1145/2601412
- Aral S., Walker D., (2013). *Tie Strength, Embeddedness & Social Influence: Evidence from a Large Scale Networked Experiment* (SSRN Scholarly Paper No. ID 2197972). Social Science Research Network, Rochester, NY.
- Arnaboldi V., Conti M., Passarella A., Dunbar, R. (2013). Dynamics of Personal Social Relationships in Online Social Networks: A Study on Twitter, *Proceedings of the First ACM Conference on Online Social Networks, COSN'13*. ACM, New York, NY, USA, p. 15-26. doi:10.1145/2512938.2512949
- Bennett, P.N., White, R.W., Chu, W., Dumais, S.T., Bailey, P., Borisyuk, F., Cui, X. (2012). Modeling the Impact of Short- and Long-term Behavior on Search Personalization, in: *Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR'12*. ACM, New York, NY, USA, p. 185-194. doi:10.1145/2348283.2348312
- Cabanac G. (2011). Accuracy of inter-researcher similarity measures based on topical and social clues. *Scientometrics*, 87, p. 597-620. doi:10.1007/s11192-011-0358-1
- Canut C.M.-F., On-at S., Péninou A., Sèdes F. (2015). Enrichissement du profil utilisateur à partir de son réseau social dans un contexte dynamique : application d'une méthode de pondération temporelle, in: *Actes du XXXIII^e Congrès Inforsid*, Biarritz, France, May 26-29, p. 15-30.
- Carmel D., Zwerdling N., Guy I., Ofek-Koifman S., Har'el N., Ronen I., Uziel E., Yogev S., Chernov S. (2009). Personalized Social Search Based on the User's Social Network, in: *Proceedings of the 18th ACM Conference on Information and Knowledge Management, CIKM '09*. ACM, New York, NY, USA, p. 1227-1236. doi:10.1145/1645953.1646109
- Cazabet R., Leguistin M., Amblard F. (2012). Automated Community Detection on Social Networks: Useful? Efficient? Asking the Users, in: *Proceedings of the 4th International Workshop on Web Intelligence; Communities*. ACM, New York, NY, USA, p. 6:1-6:8. doi:10.1145/2189736.2189745
- Crabtree B., Soltysiak S., Pp M., Re I. (1998). Identifying and tracking changing interests. *Int. J. Digit. Libr.* 2, p. 38-53.
- Ding, Y., Li X. (2005). Time Weight Collaborative Filtering, in: *Proceedings of the 14th ACM International Conference on Information and Knowledge Management, CIKM'05*. ACM, New York, NY, USA, p. 485-492. doi:10.1145/1099554.1099689
- Gama J., Žliobaite I., Bifet A., Pechenizkiy M., Bouchachia A. (2014). A Survey on Concept Drift Adaptation. *ACM Comput Surv*, vol. 46, n° 4, p. 1-44. doi:10.1145/2523813

- Gauch S., Speretta M., Chandramouli A., Micarelli A. (2007). User Profiles for Personalized Information Access, *The Adaptive Web, Lecture Notes in Computer Science*. Brusilovsky P., Kobsa A., Nejdl W. (Eds.), Springer Berlin Heidelberg, p. 54-89.
- Gomez Rodriguez M., Leskovec J., Schölkopf B. (2013). Structure and Dynamics of Information Pathways in Online Media, in: *Proceedings of the Sixth ACM International Conference on Web Search and Data Mining, WSDM'13*. ACM, New York, NY, USA, p. 23-32. doi:10.1145/2433396.2433402
- Jiang C., Chen Y., Liu K.J.R. (2014). Evolutionary Dynamics of Information Diffusion Over Social Networks. *IEEE Trans. Signal Process.* 62, p. 4573-4586. doi:10.1109/TSP.2014.2339799
- Kacem A., Boughanem M., Faiz R. (2014). Time-Sensitive User Profile for Optimizing Search Personalization, *User Modeling, Adaptation, and Personalization, Lecture Notes in Computer Science*, Dimitrova V., Kuflik T., Chin D., Ricci F., Dolog P., Houben G.-J. (Eds.), Springer International Publishing, p. 111-121.
- Koren Y. (2009). Collaborative Filtering with Temporal Dynamics, *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD'09*. ACM, New York, NY, USA, p. 447-456. doi:10.1145/1557019.1557072
- Kumar R., Novak J., Tomkins A. (2006). Structure and Evolution of Online Social Networks, in: *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD'06*. ACM, New York, NY, USA, p. 611-617. doi:10.1145/1150402.1150476
- Leskovec J., Huttenlocher D., Kleinberg J. (2010). Predicting Positive and Negative Links in Online Social Networks, in: *Proceedings of the 19th International Conference on World Wide Web, WWW'10*. ACM, New York, NY, USA, p. 641-650. doi:10.1145/1772690.1772756
- Liben-Nowel, D., Kleinberg J. (2003). The Link Prediction Problem for Social Networks, in: *Proceedings of the Twelfth International Conference on Information and Knowledge Management, CIKM'03*. ACM, New York, NY, USA, p. 556-559. doi:10.1145/956863.956972
- Li, D., Cao P., Guo Y., Lei M. (2013). Time Weight Update Model Based on the Memory Principle in Collaborative Filtering. *J. Comput.* 8, p. 2763-2767. doi:10.4304/jcp.8.11.2763-2767
- Li L., Zheng L., Yang F., Li T. (2014). Modeling and broadening temporal user interest in personalized news recommendation. *Expert Syst. Appl.* 41, p. 3168-3177. doi:10.1016/j.eswa.2013.11.020
- Lin Y.-R., Chi, Y., Zhu S., Sundaram H., Tseng B.L. (2009). Analyzing Communities and Their Evolutions in Dynamic Social Networks. *ACM Trans Knowl Discov Data* 3, 8:1-8:31. doi:10.1145/1514888.1514891
- Maloof M.A., Michalski R.S. (2000). Selecting Examples for Partial Memory Learning. *Mach. Learn.* 41, p. 27-52. doi:10.1023/A:1007661119649
- Mezghani M., On-at S., Péninou A., Canut M.-F., Zayani C.A., Amous I., Sedes F. (2015). A Case Study on the Influence of the User Profile Enrichment on Buzz Propagation in

- Social Media: Experiments on Delicious, *New Trends in Databases and Information Systems, Communications in Computer and Information Science*. Morzy T., Valduriez P., Bellatreche L. (Eds.), Springer International Publishing, p. 567-577.
- Mezghani M., Zayani C.A., Amous I., Gargouri F. (2012). A User Profile Modelling Using Social Annotations: A Survey, in: *Proceedings of the 21st International Conference Companion on World Wide Web, WWW'12 Companion*. ACM, New York, NY, USA, p. 969-976. doi:10.1145/2187980.2188230
- Mianowska B., Nguyen N.T. (2013). Tuning User Profiles Based on Analyzing Dynamic Preference in Document Retrieval Systems. *Multimed. Tools Appl* 65, p. 93-118. doi:10.1007/s11042-012-1145-6
- Rivera M.T., Soderstrom S.B., Uzzi B. (2010). Dynamics of Dyads in Social Networks: Assortative, Relational, and Proximity Mechanisms. *Annu. Rev. Sociol.* 36, p. 91-115. doi:10.1146/annurev.soc.34.040507.134743
- Rodriguez M.G., Leskovec, J., Schölkopf B. (2012). Structure and Dynamics of Information Pathways in Online Media. ArXiv12121464 Phys.
- Shaw J.A., Fox E.A., Shaw J.A., Fox E.A. (1994). Combination of Multiple Searches, in: *The Second Text REtrieval Conference (TREC-2)*, p. 243-252.
- Stanley, K.O. (2003). Learning Concept Drift with a Committee of Decision Trees (No. AI03-302). Department of Computer Sciences, The University of Texas at Austin.
- Stattner E., Collard M., Vidot N. (2013). D2SNet: Dynamics of diffusion and dynamic human behaviour in social networks. *Comput. Hum. Behav., Advanced Human-Computer Interaction* 29, p. 496-509. doi:10.1016/j.chb.2012.06.004
- Sugiyama K., Hatano K., Yoshikawa M. (2004). Adaptive Web Search Based on User Profile Constructed Without Any Effort from Users, in: *Proceedings of the 13th International Conference on World Wide Web, WWW'04*. ACM, New York, NY, USA, p. 675-684. doi:10.1145/988672.988764
- Tan B., Shen X., Zhai C. (2006). Mining Long-term Search History to Improve Search Accuracy, in: *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD'06*. ACM, New York, NY, USA, p. 718-723. doi:10.1145/1150402.1150493
- Tchuente, D. (2013). Modélisation et dérivation de profils utilisateurs à partir de réseaux sociaux : approche à partir de communautés de réseaux k-égocentriques (phd). Université de Toulouse, Université Toulouse III - Paul Sabatier.
- Tchuente D., Canut M.-F., Jessel N., Peninou A., Sèdes F. (2013). A community-based algorithm for deriving users' profiles from egocentrics networks: experiment on Facebook and DBLP. *Soc. Netw. Anal. Min.* 3, p. 667-683. doi:10.1007/s13278-013-0113-0
- Tsymbol A. (2004). *The Problem of Concept Drift: Definitions and Related Work, Technical report*. Department of Computer Science, Trinity College: Dublin.
- Tylenda T., Angelova R., Bedathur S. (2009). Towards Time-aware Link Prediction in Evolving Social Networks, in: *Proceedings of the 3rd Workshop on Social Network*

Mining and Analysis, SNA-KDD'09. ACM, New York, NY, USA, p. 9:1–9:10.
doi:10.1145/1731011.1731020

Weng L., Ratkiewicz J., Perra N., Gonçalves B., Castillo C., Bonchi F., Schifanella R., Menczer F., Flammini A. (2013). The Role of Information Diffusion in the Evolution of Social Networks, in: *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD'13*. ACM, New York, NY, USA, p. 356-364. doi:10.1145/2487575.2487607

Widmer G., Kubat M. (1993). Effective learning in dynamic environments by explicit context tracking, *Machine Learning: ECML-93*, Lecture Notes in Computer Science. Brazdil, P.B. (Ed.), Springer Berlin Heidelberg, p. 227-243.